ABSTRACT

Disclosed are methods for identifying nucleic acid sequences which are of different abundances in different nucleic acid source populations, e.g. differentially expressed genes or genomic variations among individuals or populations of individuals. In one embodiment, probes derived from the source nucleic acid populations are derivatized with a terminal sample ID (SID) sequence characteristic of that population. Upon competitive hybridization of the probes to a reference or index nucleic acid library containing all the sequences in the populations being compared, the SID tags remain single stranded, and those from the different sources are then annealed to one another. Unhybridized (remainder) SID sequences are then quantified. By labeling such remainder SID sequences with a fluorescent dye, FACS sorting of beads containing the hybridized probes can be carried out. The signal ratio upon which such sorting is based is enhanced compared to competitive hybridization using labeled probes without SID sequences.